NANO LETTERS

2006
Vol. 6, No. 7
1483−1486

Sequencing DNA by Dynamic Force Spectroscopy: Limitations and Prospects

N. K. Voulgarakis,* A. Redondo, A. R. Bishop, and K. Ø. Rasmussen

Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received April 19, 2006; Revised Manuscript Received May 22, 2006

ABSTRACT

Using a well established model, we systematically analyze fundamental limitations on the viability of using mechanical unzipping of DNA as a fast and inexpensive sequencing method. Standard unzipping techniques, where double-stranded DNA is unzipped through the application of a force at one end of the molecule, are shown to be inadequate. Emerging techniques that unzip DNA by local force application are more promising, and we establish the necessary experimental requirements that must be met for these techniques to succeed as single molecule sequencing tools.

Development of fast and accurate DNA sequencing methods is certain to spur significant advances in biomedical research and biotechnology. Important examples are the development of individualized drugs and preventive measures against various common diseases.[1] Additional to the obvious biomedical interest, DNA sequencing may also lead to important achievements in computer science. Since more than 10^{21} bits of information can be stored in 1 g of dehydrated DNA, it has been suggested as a data storage medium with orders of magnitude higher capacity than current devices.[2] Similarly, DNA-based computing could dramatically increase the efficiency of numerical simulations currently provided by silicon-based computers.[3] Current DNA sequencing techniques are, however, very expensive, time-consuming, and far too inaccurate for such broad applications. Therefore, it is clearly desirable to develop a next generation of sequencing tools, enabling rapid, massive sequencing at the single molecule level.

Single molecule mechanical unzipping of double-stranded DNA has been proposed as a possible sequencing method, since the binding strengths between complementary base pairs can potentially be accurately detected by a force probe during the unzipping process. If the experimental apparatus can achieve sufficient sensitivity to probe a single base pair unzipping event, then the sequence information can possibly be extracted from the force signal. In such a case, DNA sequencing can be done in a matter of hours with a nonprohibitive cost for general use. Several different experimental setups[4] have been suggested for the mechanical unzipping of DNA and the best resolutions have been achieved by optical tweezers (∼100 base pair resolution)[5] and recently by atomic force microscopy (AFM) (∼10 base pair resolution), but this technique is limited to small molecules.[6]

The aim of this work is to establish theoretical limits on the applicability of mechanical unzipping as an efficient sequencing tool. First, we will analyze fundamental restrictions that plague current experimental techniques. Second, we will investigate whether novel local unzipping techniques under current development can sufficiently improve resolution and accuracy. We will use a simple one-dimensional model introduced by Peyrard−Bishop−Dauxois (PBD)[7,8] that has been demonstrated to accurately predict details of the thermal denaturation as well as the sequence dependence of local openings (bubbles) in DNA.[9−11] Very recent studies have additionally demonstrated that the predictions of the PBD model are also in excellent agreement with experimental measurements of the mechanical unzipping of DNA.[12]

A schematic representation of our modeling framework is shown in Figure 1. The dynamics of double-stranded DNA are described by a single degree of freedom corresponding to the relative distortion of a base pair from its equilibrium position. The hydrogen bonds between the complementary bases are phenomenologically described by Morse potentials, while the interactions between nearest base pairs are represented by nonlinear springs.[8,9] Similarly to a standard unzipping experiment, we consider the first base to be fixed to a substrate, while its complementary base is tethered to a force probe through a polymeric link. For the sake of simplicity, the total stiffness of the link and the force probe is modeled by a linear spring of stiffness k_0. The double strand is gradually unzipped by moving the force probe at

10.1021/nl0608801 CCC: $33.50
Published on Web 06/03/2006

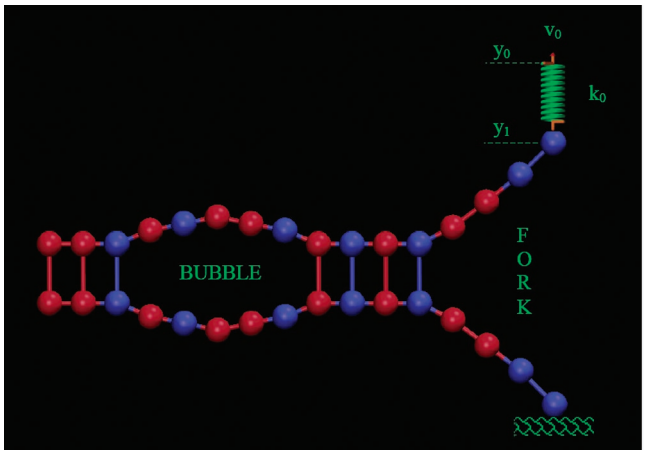

Figure 1. A schematic representation of the PBD model in a dynamic force spectroscopy experiment.

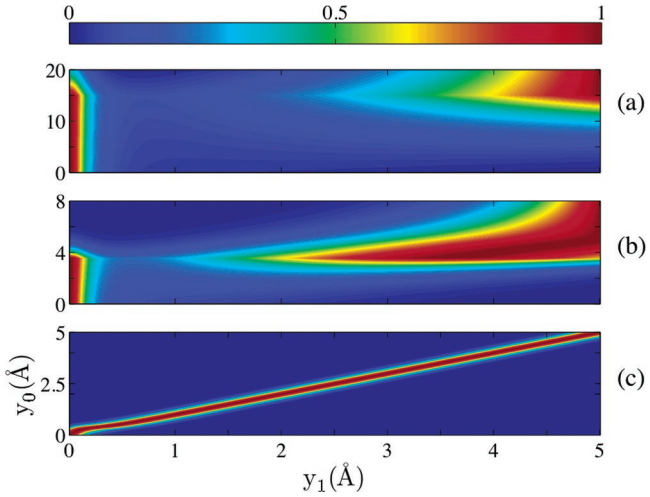

Figure 2. The density plot of the probability $P(y_1,y_0)$ of a single bond for (a) $k_0 = 16$, (b) $k_0 = 160$, and (c) $k_0 = 1.6 \times 10^4$ pN/nm. In all cases the temperature is 300 K (see text for details).

constant velocity. The force needed to unzip the molecule at a given extension is determined using a simple Monte Carlo method.[12] More specifically, we displace the force probe by a distance Δy ($y_0 \rightarrow y_0 + \Delta y$) and then perform N Monte Carlo steps to compute the resulting average distortion of the molecule. The force is then given by Hooke's law: $F = k_0(y_0 - \langle y_1 \rangle)$, where $\langle y_1 \rangle$ is the position of the first base pair. Within this numerical approach, the speed of pulling is defined as $v_0 = \Delta y/N$ (see ref 12 for more details). The sequence of the molecule is recorded in the force extension profile; adenine−thymine (AT) rich regions are unzipped by a force of 9−10 pN, while guanine−cytosine (GC) rich regions require a force of 15−20 pN, which is in excellent agreement with experimental observations.[5,12−14]

Although such dynamical force spectroscopy methods can provide very important information on the stability of the molecule, it is difficult to use it for DNA sequencing. The difficulty can be directly attributed to the stiffness of the experimental force probe apparatus, as well as to the internal flexibility of the molecule, which both impose severe restrictions on the applicability of this technique as a sequencing tool.

The dissociation of a single bond provides a suitable illustration of these limitations.[15] We consider that the bond is well approximated by a Morse potential, $V(y_1) = D[1 - \exp(-ay_1)]^2$, where y_1 is the distortion from the equilibrium position, while $D = 0.063$ eV and $a = 5.5$ Å^{-1}. These parameter values are typical for the description of a base pair interaction.[9] Within this description the bond can be considered as bound if $y_1 < 0.5$ Å and unbound beyond this point. The experimentally detectable areas of this bonding potential are determined by the equilibrium probability $P(y_1,y_0) \propto \exp[-\{2V(y_1) + k_0(y_1 - y_0)^2\}/2kT]$ of finding the bond with distortion y_1, while the force probe is at a distance y_0 away from the initial position. The larger this probability, the more easily the state can be detected by the force probe. In Figure 2 we show the probability $P(y_1,y_0)$ at room temperature, for three different probe stiffnesses. For relative small values of k_0 (<160 pN/nm) (plots a and b), there is a significant barrier between the bound and the unbound state of the bond. States close to the barrier, where the probability vanishes, are experimentally inaccessible. However, this inaccessible area decreases with increasing probe stiffness and for $k_0 > 1.6 \times 10^4$ pN/nm (plot c), the dissociation process becomes continuous with no inaccessible regions.[18] Consequently we see that the spatial resolution can be improved by stiffer probes.

In the case of DNA unzipping, an ever increasing amount of flexible single strand is created between the force probe and the unzipping site. This results in a problematic decrease in the combined stiffness of the probe and the additional single strand. The single strand will quickly dominate the effective stiffness of the apparatus, and the resolution consequently deteriorates. An effective way of illustrating the resolution deterioration of a dynamics force spectroscopy experiment is to study periodic DNA molecules composed of repeating AT and GC blocks, each containing N base pairs. Following the experimental work of ref 6, we determine the number of base pairs that can be detected based on the force-extension profiles. In Figure 3a, we plot the force extension curves for $N = 10$, 20, and 30 at room temperature. The periodicity of the molecule is reflected in the force extension profile, but in all three cases the force signal decreases with increasing extension and vanishes completely when the effective stiffness becomes too small. In this regime, the signal becomes flat and the force obtained merely corresponds to the mean value of the force needed to unzip AT and GC base pair. However, Figure 3 clearly demonstrates that the larger N becomes, the more repeating blocks the method can detect.[19] At room temperature we were unable to obtain resolutions better than $N = 5$. The resolution of this technique can be improved by reducing thermal fluctuations. This is illustrated in Figure 3b, where we present the force−extension curve of a periodic molecule ($N = 10$), for three different temperatures. As shown, the force signal variation becomes more distinct and persistent as the temperature decreases. However the signal eventually vanishes even at very low temperatures. Even, in the practically unachievable, zero temperature limit, sequencing of large molecules remains impossible: With the thermal fluctuations

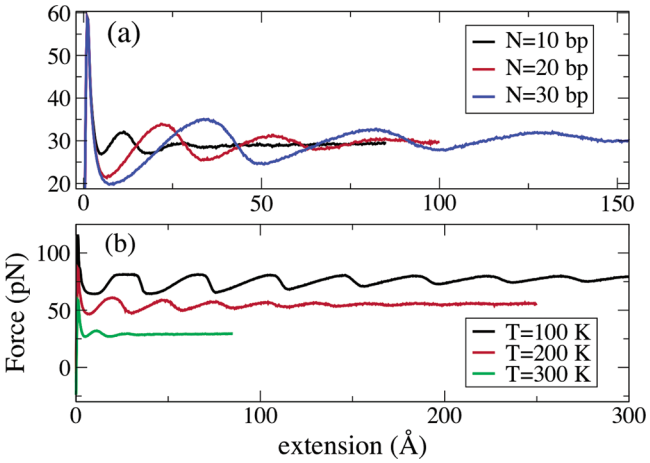

Figure 3. Force−extension profile of (a) periodic sequences with $N = 10$, 20, and 30 at room temperature, and (b) periodic sequence with $N = 10$ at $T = 300$, 200, and 100 K. Each curve is the average over 10^3 independent simulations. In all three cases the pulling speed is $v_0 = 10^{-5}$ Å/MC steps, the probe stiffness $k_0 = 1600$ pN/nm, while the total length of the molecule is 300 base pairs.

neglected, the single-strand stiffness drops as $k/2j$, where k is the stacking interaction strength and j is the number of the unzipped base pairs. With $k = 400$ pN/nm[9] and $k_0 = 100$ pN/nm, which is the stiffness of a typical AFM cantilever,[4] the effective stiffness for $j = 100$ base pairs is lower than 4 pN/nm. This is insufficient[12] for the detection of a single base pair, so even at zero temperature, sequencing of large molecules remains impossible.

The development of experimental techniques that circumvent the problems caused by the increasing single-strand link between the force probe and unzipping sites is currently underway. Such methods attempt to apply a principle similar to a familiar clothing zipper by always applying the force locally.[16] The basic concept is to attach a nanoscopic ring (realized for instance by a cyclodextrin molecule or small nanotubes) to an AFM cantilever and thread one of the DNA strands through the ring. As the force probe moves, the single strand slides through the ring unzipping the double-stranded molecule. The forces applied through this procedure are then detected by the deflection of the AFM cantilever. Although, the accuracy of this technique of course still depends on the effective stiffness of the AFM cantilever and the molecular ring, the advantage is that the force probe stiffness remains constant throughout the entire process.

To simulate this process within the PBD model, we approximate the effective stiffness of the AFM cantilever and the molecular ring by a linear spring. Once a given base pair is unzipped, then the force probe relaxes and moves to the next base pair. This process is repeated until the final base pair of the molecule is unzipped. In Figure 4a, we present the force detected by the probe as a function of time for a periodic molecule ($N = 5$) at room temperature. This saw-tooth structure can be interpreted by relating the higher rupture forces to the stronger GC base pairs and the lower forces to the weaker AT base pairs. However, as shown, the molecular sequence cannot be uniquely obtained by a single simulation. Some unzipping events that correspond to GC

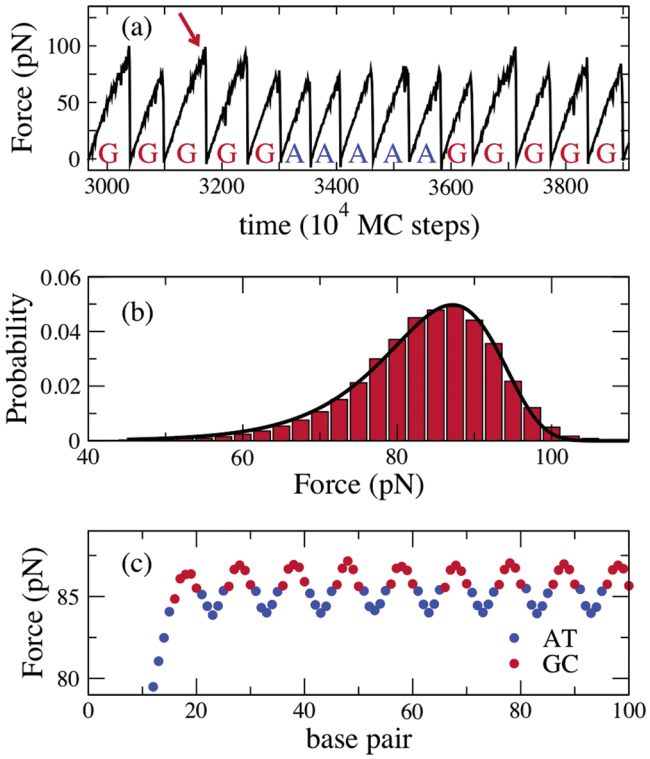

Figure 4. Unzipping DNA by a local force probe. (a) Force−time profile of the unzipping a periodic DNA with $N = 5$ by a local force probe. (b) Probability density function of the unzipping event indicated by the arrow in (a). (c) Most probable rapture force versus base pair. The force probe stiffness is $k_0 = 160$ pN/nm, and $v_0 = 10^4$ Å/MC steps.

base pairs require lower rupture forces than AT. This phenomenon is attributed to the stochastic nature of the process; there is always a finite probability for a GC base pair to rupture at a lower force than a AT base pair. More accurate detection of the base pair variation can be achieved by obtaining the most probable force of a given unzipping event. The probability distribution functions (PDF) for the rupture force of the unzipping event indicated by an arrow in Figure 4a is presented in Figure 4b. The most probable unzipping force (rupture force) is obtained by fitting the theoretical PDF for a single bond dissociation under a constant force loading rate[15] to the numerical data. In Figure 4c, we present the most probable rupture force as a function of base pairs for a $N = 5$ sequence and the sequence periodicity can now be accurately extracted from this profile. Additionally, we observe (directly int the simulations) that the base pairs tend to rupture in bursts rather than one by one. If the method could detect single base pair rupture events, the obtained force signal would be a step function; the lower (higher) value corresponding to the rupture of a AT (GC) base pairs. Instead we observe a sinusoidal force signal indicating that each unzipping event is affected by the neighboring base pairs. We can, for instance, see that the AT base pairs closer to GC blocks require higher force to unzip than those in the middle of an AT block. The uncertainty of 10−15 base pairs at the beginning of the process is caused by the open boundary conditions. This problem disappears once the force probe location is beyond

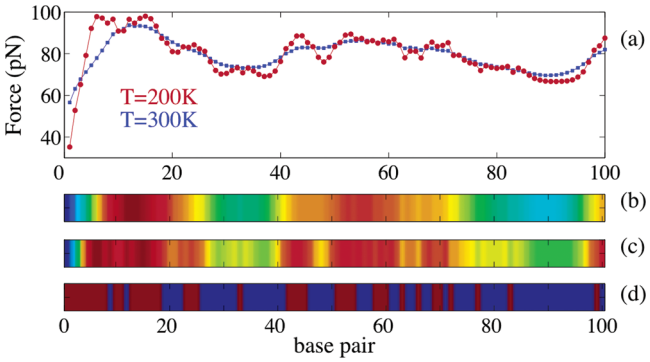

Figure 5. (a) Force vs base pair for the first 100 base pairs of the λ-DNA at $T = 300$ and 200 K. The curve of 200 K is shifted down to the mean value of that at 300 K, for ease of comparison. (b, c) The density plots of the curves presented in (a) for $T = 300$ K and $T = 200$ K, respectively. (d) The ideal density plot for the real sequence (red region correspond to GC base pair, and blue to AT). In all cases $k_0 = 160$ pN/nm, and $v_0 = 10^4$ Å/MC steps.

$\simeq$15 base pairs, because then the double-stranded DNA recombine behind the ring. The resolution of this technique is indeed better than that of the more standard unzipping experiments discussed earlier. An important final observation is that our numerical simulations indicate that chancing the pulling speed does not significantly affect the force resolution. This is caused by the resolution-necessitated use of a rather stiff force probe. As explained by Figure 2, the energy barrier between bound and unbound states is not significant in this case and the rupture force is consequently relatively independent of the pulling speed.

A more challenging test of this method is the sequencing of a real DNA molecule. In Figure 5a we present the force signal obtained for the first 100 base pairs of λ-DNA at $T = 300$ K (blue line) and $T = 200$ K (red line), respectively. Although, there is a clear sequence-dependent structure, we were unable to retrieve the molecular sequence by following the procedure used in the previous paragraph. This is further illustrated, in parts b and c of Figure 5, where we present the density representation of the data presented in (a) at 300 and 200 K, respectively together with a color representation of the sequence (GC is shown in red and AT in blue) (see Figure 5d). We see that lower temperature signals are better correlated with the actual sequence. This is because, at low temperatures, the density and size of the spontaneously formed bubbles are significantly reduced, and this results in a lower uncertainty in DNA sequencing. The development of an effective filtering method, based on the result of known sequences, could possibly provide an adequate estimation of the real sequence of the molecule.[17]

In summary, we have presented a numerical analysis of DNA sequencing by dynamic force spectroscopy. In particular, we have shown that current experimental techniques for DNA unzipping cannot be used as efficient sequencing tools. Even in the case of a perfect force sensor and at sufficiently low temperatures, these techniques are inadequate to provide reliable data, because of the high flexibility of the single-stranded DNA. We have also investigated the resolution and reliability of a more sophisticated method, under development, used for local unzipping of DNA. We conclude that this method could provide a source of DNA sequencing in the case of (a) very stiff and accurate force probes, (b) a large number of measurements being performed for each molecule, and (c) the development of an effective filtering method for the experimental data acquired. Further improvement can be achieved by decreasing the thermal fluctuations in a lower temperature regime.

Acknowledgment. We thank Professor S. M. Lindsay for fruitful discussions. Work at Los Alamos is supported by the U.S. Department of Energy.

References

(1) Shendure, J.; Mitra, R. D.; Varma, C.; Church, G. M. *Nat. Rev. Genet.* **2004**, *5*, 335.
(2) Reif, J. H. *Science* **2002**, *296*, 478.
(3) Adleman, L. M. *Science* **1994**, *266*, 1021.
(4) Conroy, R. S.; Danilowicz, C. *Contemp. Phys.* **2004**, *45* (4), 277.
(5) Rief, M.; Clausen-Schaumann, H.; Gaub, H. E. *Nat. Struct. Biol.* **1999**, *6*, 346.
(6) Krautbauer, R.; Rief, M.; Gaub, H. E. *Nano Lett.* **2003**, *3* (4), 493.
(7) Peyrard, M.; Bishop, A. R. *Phys. Rev. Lett.* **1989**, *62*, 2755.
(8) Dauxois, T.; Peyrard, M.; Bishop, A. R. *Phys. Rev. E* **1993**, *47*, R44.
(9) Campa, A.; Giansanti, A. *Phys. Rev. E* **1998**, *58*, 3538.
(10) Voulgarakis, N. K.; Kalosakas, K.; Rasmussen, K. Ø.; Bishop, A. R. *Nano Lett.* **2004**, *4* (4), 629.
(11) Ares, S.; Voulgarakis, N. K.; Rasmussen, K. Ø.; Bishop, A. R. *Phys Rev. Lett.* **2005**, *94*, 035504.
(12) Voulgarakis, N. K.; Redondo, A.; Bishop, A. R.; Rasmussen, K. Ø. *cond-mat/0512487* **2006**.
(13) Essevaz-Roulet, B.; Bockelmann, U.; Heslot, F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11935.
(14) Danilowicz, C.; Kafri, Y.; Conroy, R. S.; Coljee, V. W.; Weeks, J.; Prentiss, M. *Phys. Rev. Lett.* **2004**, *93*, 078101.
(15) Evans, E.; Ritchie, K. *Biophys J* **1999**, *76*, 2439.
(16) Lindsay, S. M. Private communication.
(17) Voulgarakis, N. K.; Redondo, A.; Bishop, A. R.; Rasmussen, K. Ø. In preparation.
(18) In this limit, the elastic energy of the probe dominates the total energy and, as a result, there is only a single minimum.
(19) The elevated forces observed in this simulation compared to the experimental forces are a result of our use of a stiffer probe in order to increase the resolution.[15]

NL0608801